Multimedia learning: beyond modality

P. Reimann *

University of Heidelberg, Institute of Psychology, Hauptstrasse 47–51, D-69120 Heidelberg, Germany

The technical means to develop informational and instructional resources are now at everybody’s fingertips (e.g., by using MS PowerPoint, the most widely used multimedia and instructional “authoring tool”, although by design only a presentation (sic!) tool). It also has become very easy to distribute multimedia materials: one does not need to press CD-ROMs and ship them via mail, all that is needed is to upload resources to a web server; modern browsers can render all kinds of media, from text and pictures to animated 3D models. Having the hurdle of technical and logistical problems moved out of the way, issues of learning and didactics become more prominent again: how does multimedia learning work? Under what conditions does it help to present content with multiple media? How can students become engaged in active learning, in interacting with media? What are the well-founded (general) design principles? The contributions in this Special Issue address these and a number of related questions in great depth and with scientific rigor.

All of the papers make important contributions not only to current issues in multimedia learning research, but also result in insights that are of relevance for practical applications—for designing multimedia-based instructional messages and learning environments. It is notoriously difficult to write for both purposes in a single paper—to foster research and to contribute to instructional practice. Most of the contributions speak primarily to the researcher. Because I find it important that readers with an interest in instructional design can profit from this Special Issue as well, one objective for this commentary is to identify and summarize the instructional messages contained in the research papers. My second objective is to try to identify central theoretical issues and to speculate upon further research directions. In order to reach these objectives, I will first step through the individual contributions and state what I see as their central points. In doing so, I shall not follow the exact sequence of articles.

Of all the papers, the one by Richard Mayer is most clearly directed towards the
instructional designer community. In summarizing concisely his and his co-workers’ research on the combination of (written and spoken) words with (static and dynamic) pictures, Mayer identifies four effects that hold in book-based as well as computer-based learning environments: (a) a multimedia effect—students learn more deeply from words and pictures than from words alone; (b) a coherence effect—students learn more deeply when extraneous material is excluded; (c) a spatial contiguity effect—deeper learning ensues when printed words are placed near corresponding pictures; and (d) a personalization effect—conversational style leads to deeper processing than a more formal style. These effects were established in the context of numerous experimental studies where students had to learn about simple mechanical systems such as brakes and pumps. “Depth” of learning (or understanding) was assessed by problem-solving transfer, performance in troubleshooting tasks for instance. As the effects are phrased in terms of the characteristics of multimedia presentations, they are of immediate value for designers who can turn these effects into design principles.

Most of the effects—certainly (a) to (c)—can be explained by means of dual coding theory (Clark & Paivio, 1991; Paivio, 1986) combined with an element or two from cognitive load theory (Chandler & Sweller, 1991). Mayer, in elaborating dual coding theory, assumes that verbal and pictorial information is processed in two different cognitive subsystems, leading to a word-based representation and a picture-based representation. These two separate representations are then integrated by means of a process that produces mappings between verbal and pictorial elements and relations. The advantage of such an integrated representation is that it is indexed in memory in multiple formats and hence can be accessed in more than one way. Mayer’s fourth effect, the “personalization” effect, seems not to be directly related to this memory theory. Mayer does not elaborate this point, but we can speculate that it is either a motivational effect and/or maybe related to the advantages of narrative formats for retention (e.g., Galambos, Abelson, & Black, 1986).

Wolfgang Schnotz and Maria Bannert, in the next article, argue that dual coding theory—also in the elaborated form provided by Mayer (1997)—is not sufficient as an explanatory account for the multimedia effect. They claim that the mapping process between the verbal and the pictorial representation formats as foreseen in dual coding theory cannot be conceptualized as a structural integration because verbal (or propositional) representations are descriptive whereas pictorial (or analogue) representations are depictive. A descriptive representation consists of arbitrary, convention-based symbols both for elements and relations. Depictive representations, on the other hand, consist of iconic signs and do not contain signs for relations. Hence, a simple structure mapping between descriptive and depictive representations is not possible; in particular, relations cannot be mapped as there are none in the depictive relation.

Schnotz and Bannert suggest a new model for the integration of information in word and pictures that is different in a number of ways from dual coding theory. The main differences are these. Firstly, dual coding is assumed not only for the processing of pictures, but also for the processing of words and texts. Secondly, the interaction between verbal and pictorial processing does not primarily take place on
the level of text surface representation and visual perception/image, but on the level of propositional representation and mental models, i.e., after the semantic processing of both text and picture information. Thirdly, what is seen in dual coding theory as a structurally integrated representation of verbal and pictorial representation is in the new theory seen as a continuous process of mental-model construction and inspection; descriptive and depictive representations cannot be mapped on to each other, but they can “communicate” with each other. These two representational systems are cognitive “modules”. Seeing them as interacting modules rather than as an integrated memory structure has not only the advantage that by this we can account for the difference between descriptive and depictive representations. It can also explain why dual coding makes sense evolutionally: descriptive and depictive representations complement each other, for instance with regard to speed–accuracy tradeoffs.

Schnotz and Bannert give in their experimental study an example for such a complementary relation.

Schnotz and Bannert’s contribution is mainly an important theoretical (and empirical) one; immediate consequences for the design of multimedia instruction are not identified. One instructional issue that is discussed is one of differential indication; some findings indicate that learners with high prior knowledge when given pictures in addition to text are sometimes hampered in their knowledge acquisition process; only learners with a low degree of prior knowledge profit in general from additional pictures, provided the pictures are well designed and integrated with the text. The hampering effect can be explained by the Schnotz/Bannert model: learners with a high degree of prior knowledge can construct a mental model on their own based on the text alone. If additional pictures are provided, these may interfere with the self-generated mental models.

Tina Seufert builds on the model developed by Schnotz and Bannert and asks a question that is instructionally motivated: does help provided for learners increase coherence formation, i.e., is the connection between text and pictures increased? Which kind of help is better, if at all: directive or non-directive? And how does the amount of prior knowledge mediate the effects of help? Non-directive help is provided by general hints that relations between text passages and pictures exist, but the details are left open. Directive help provides explicit hints concerning the relevance of elements and relations within verbal and pictorial information and concerning the mappings that can be constructed. Seufert’s findings indicate that directive help as provided in this study increases both retention and comprehension whereas non-directive help does not. This is explained with the high demands that successful utilization of non-directive help imposes on learners. Furthermore, she observed that only subjects with a medium level of prior knowledge profit from the directive help provided. Less knowledgeable subjects would require more domain knowledge in order to utilize the help information. For the group with high prior knowledge, it is not clear if they did not profit from help because the help information interfered with their own processing or if they suffered from a kind of “illusion of knowledge” effect, as their actual performance was far from optimal.

Richard Lowe argues that animations are problematic for learning, in particular for early learning when little domain knowledge is developed. He argues that this
might be due to an intra-representation split-attention effect analogous to the inter-representation effect described for picture–text combinations: when learning from animations, relations have to be constructed among the different frames of the animation and this might tax the capacities of novices too much. In addition to this cognitive load problem, he observes in his study an additional problem: when studying animated weather maps (the domain under study being meteorology), learners with comparatively little domain knowledge perceive and remember mostly those characteristics of the display that are perceptually salient, i.e., components of the animation that change either substantially more than their surroundings or substantially less than their surroundings. However, for a full interpretation of weather maps, characteristics have to be taken into account that are perceptually more subtle than those readily extracted by the subjects. Lowe’s finding that novices fall prey to the fact that visual salience is not in line with conceptual significance is of course analogous to other findings in expertise research, for instance in physics (Chi, Feltovich, & Glaser, 1981) or chemistry (Kozma & Russell, 1997). The general instructional implication can only be that animations need to be very carefully designed when presenting them to novices as a learning resource. In particular, animation designers have to ensure that the visually salient features of an animation are as much as possible aligned with the conceptually important components.

That animations are a more problematic learning resource than instructional designers are often aware of is also confirmed by Doris Lewalter’s study. She analyzed the kind of learning strategies employed (spontaneously) by students when studying static and dynamic displays of a complex physical phenomenon. She found in her data that dynamic displays do not lead to more learning gains compared to—carefully designed—static displays. Based on an analysis of thinking aloud protocols, she further found that subjects employed strategic reasoning infrequently and used only simple strategy types, repetitions mostly. If strategic reasoning is employed at all, it only shows among the students who studied the static displays. Even if we suspected that verbal data do not reveal all of the strategic control behind visual processing, the very low frequency of strategic reasoning and the fact that almost none of it occurred when processing animations are clear indicators that, from an instructional perspective, learners cannot be simply left to their own devices when processing visual information, but must be systematically and extensively supported.

The final two contributions see multimedia not only as a way to present information, but also as a tool to develop external representations. Stern, Aprea, and Ebner compare in two studies groups that passively encounter a linear graph with groups that actively construct graphs from text. The rationale behind this comparison is that learners who actively graph should become more aware of the tool nature of graphs which in turn should result in better transfer; learners who see graphs more as displays than tools should exhibit less transfer. By assessing formal mathematical competencies and domain-specific knowledge, possible differential effects are taken into account. The outcomes of the two studies reported in this contribution support the central hypothesis: active graphing leads to better transfer. This finding is in line with the hypothesis that these learners become more aware of the representational
elements of linear graphs and become more aware of the possible mappings between
these representational elements and problem information.

Robert Kozma also stresses the perspective of “external representations as tools”,
but goes beyond an analysis of the purely individual usage by studying how different
representational devices are used when working with others. Earlier experimental
work (Kozma & Russell, 1997) showed that experts (in chemistry) are better able
than novices to cluster external representations of chemical phenomena into concept-
tually meaningful categories and that they can better transform one representation
(for instance, a video of a reaction) into another (a verbal description or an equation,
for instance). These differences in representational competence were further studied
in more naturalistic research (Kozma, Chin, Russell, & Marx, 2000) by observing
what chemists are actually doing with different representations in their laboratories
and by comparing this to how students perform chemical experiments and reason
about chemistry in different settings (in a wet lab, in front of a computer-based
modelling tool). With respect to representational competence, the naturalistic studies
portray a picture quite comparable to the results from the experimental studies: nov-
ices seem to be more tied into a single representation at any given point in time,
whereas experts make more use of multiple representation, switching between them
seemingly without effort. The naturalistic studies provide also information on the
purpose of utilizing multiple representations: chemists use different representations
for different purposes. For example, they use structural diagrams to reason about the
composition of compounds and chemical equations to reason about the procedures
needed for synthesis. Another set of purposes arises in the context of social interac-
tion: chemists use different representations not only for reasoning and doing, but
also for communicating with other researchers. Kozma provides examples for the
role of representations in different rhetorical contexts. Furthermore, he identifies a
number of general design principles for the use of multiple representations.

Bob Kozma’s contribution is a good starting point in order to address some of
the more general issues behind current research on multimedia learning. One very
important issue is the role of external representations. From a cognitivistic perspec-
tive, the logical question to ask is how external representations are transformed into
internal ones (e.g., propositional networks, mental models) and how external rep-
resentations must be designed in order to foster knowledge acquisition. The contrib-
utions of Mayer, Schnotz and Bannert, and Lowe are exemplary (not only
prototypical) for this approach. The contributions of Seufert and Lewalter are also
in line with this “mentalistic” view as they analyze external (help information) and
internal (learning strategies) resources that support the knowledge acquisition pro-
cess. Kozma and also Stern et al. do not deny the value of this approach, but contend
that external representations can play an additional role, namely that they can be
used directly as tools for problem solving (Stern et al.), and to guide action and to
support communication (Kozma). Under this “situated” perspective, one does not
analyze so much the cognitive residues left after learning from media, but the kinds
of activities that become possible when people interact with media (learning with
media; Salomon, Perry, & Globerson, 1991). Under this more activity-oriented per-
spective, the question of how people construct external representations (e.g., graphs
as analyzed by Stern et al.) becomes as important as how they make use of representations provided by others. I find the analysis of the material features of external representations and their cognitive and social affordances a very important extension for research on multimedia learning, in particular in light of evidence that the material aspects of representations may be more important for (scientific) practice than psychologists may expect (e.g., Lynch & Woolgar, 1990; McGinn & Roth, 1999). For instance, while Lowe treats weather maps in his study primarily as a learning device, based on which learners are expected to develop their own mental representations, it would also be interesting to see how learners use—possibly interactive—weather maps as problem-solving tools.

Another important theoretical theme underlying a number of contributions is dual coding theory. Schnotz and Bannert put their fingers on some of the conceptual problems with dual coding theory as suggested by Paivio (1986) and elaborated for instructional psychology by Mayer (1997). Their modified dual coding theory includes as central new element a suggestion to solve the problem of how to relate descriptive and depictive representations. While this is theoretically sound, the new model is also substantially more complex than the original dual coding theory. Testing it empirically will require to go beyond current research. The new model is also in need of a number of clarifications. For instance, it needs to be shown that a mechanism such as structure mapping is indeed capable of producing meaningful processing results in information-rich domains. Algorithms such as structure mapping that take only syntactical features into account tend to encounter combinatorial problems. Furthermore, the model construction and inspection processes that hypothetically connect propositional representations and mental models need to be specified, preferably in computational terms so that this assumption becomes testable.

We should also keep in mind that dual coding theory—whether old or new—addresses only one dimension of multimedia: the modality, i.e., the forms of expressions that are used for displaying information. Based on it, instructional design principles can be derived for configuring picture–text combinations, but the theory is mute on a number of other important representational issues (as for instance identified in de Jong et al., 1998). For instance, decisions regarding perspective/ontology, i.e., what to represent and in what detail to represent it, are obviously quite important for instruction, but not easily answered based on dual coding theory. While such kinds of decisions may be left delegated other theories (for instance, theories on the development of everyday conceptions of scientific phenomena, curriculum theories), more worrisome is that fact that dual coding theory has nothing to say on the kind of graphical representations that one should employ in order to visualize something. This would require to be more explicit about the computational (and material, we might add) characteristics of particular depictive representations such as graphs, tables, Venn diagrams, etc. For instance, different graphical representations have different degrees of locality (Larkin and Simon, 1987), emergence (Kulpa, 1995) and inexpression (Stenning and Oberlander, 1995), all of these affecting how good a specific graphical representation fits a specific class of problem-solving tasks. Of particular importance in the context of learning is to take into account the “expressivity” of a representation format. From a computational (and instructional) point of
view, information should not be presented to learners by means of a notation (representation) that is maximally expressive, but by one that is minimally expressive (i.e., maximally specific) vis-à-vis the intended learning goals. This “specificity” principle (Stenning and Oberlander, 1995) may be instructionally as important as the question of how to configure picture–text combinations. One may also have to sequence representations with different degrees of specificity in order to enhance learning (e.g., from more to less specific), which is another area of instructional decisions where dual coding theory provides no advice.

A third issue that is treated in a number of contributions (Lowe, Mayer, Seufert) is the effects animations have on learning. While these days multimedia designers seem to follow the principle that animations should be provided as often as possible (budget allowing), the instructional value of animations is extremely difficult to establish. Animations can have detrimental cognitive effects due to the extreme cognitive load they produce and due to misalignment between conceptual relevance and perceptual salience. And this even in areas where the dynamics of a system are important to consider and hard to visualize without animations, for instance weather phenomena. We clearly do not have a satisfying theory at this point in time that could help us to decide rationally under which circumstances (domain characteristics, learner characteristics, learning goals) to use animations and how to design these animations optimally. I find cognitive load theory too unspecific to be of much help here; Lowe’s observation concerning the alignment of conceptual importance and visual salience, however, indicates that domain knowledge is again the crucial factor to take into account during design.

We find in this Special Issue some of the leading researchers in the field answering a number of questions and raising many new ones. Given that these new questions are raised in the context of rather established and well-researched forms of multimedia—text, 2D pictures, 2D animations—we can be sure that this will not be the last Special Issue on this topic. We can expect to hear in the near future more about text, pictures and animations, but hopefully also about research into three-dimensional multimedia, interactive multimedia and immersive technologies. Otherwise, we run the danger of repeating a problem encountered in man–machine interaction reasearch: once psychologists knew almost all about the psychology of the command-driven interface, there were no command-driven user interfaces in use anymore.

References


